Northeastern University, Boston, MA

College of Engineering

Department of Civil and Environmental Engineering

CIVE 7381: Transportation Demand Forecasting and Model Estimation

Problem Set 2

From: **Nathan David Obeng-Amoako** (NUID: 002607282)

To: Professor Haris Koutsopoulos

Submitted on: Friday, September 27th

Fall 2024

Table of Contents

Table of Contents ii
List of Figures iii
List of Tables iii
List of DataFrames iii
List of Equations iii
Problems iv
Solution to Problem 1 7
Problem 1a 7
Problem 1b 7
Problem 1c 7
Problem 2 8
Problem 2a 8
Problem 2b 8
Problem 2c 10
Problem 3 12
Section 3: Household Travel 12
Section 4: Person Travel 13
Section 5: Private Vehicle Travel 13
Section 6: Vehicle Use and Availability 14
Section 7: Commute Travel Patterns 15
Section 9.2: Emerging Travel Trends 16

List of Figures

Figure 1: Boxplots of the Distributions of Relevant Variables 9
Figure 2: Histograms of the Distributions of the Relevant Variables 10
Figure 3: Correlation Matrix of Relevant Variables 10
Figure 4: Scatter Matrix of Relevant Variables 11
Figure 5: Daily PMT per Person by Trip Purpose 13

List of Tables

Table 1: Average Person Trip Length by Trip Purpose 12
Table 2: Average Vehicle Occupancy for Selected Trip Purposes 14
Table 3: Number of Households by Availability of Household Vehicles 14
Table 4: Commute Patterns by Mode of Transportation 15

List of DataFrames

DataFrame 1: Descriptive Statistics for the Variables in the Dataset 9

List of Equations

Equation 1: Standard Error 7
Equation 2: Confidence Interval 7
Equation 3: Number of Samples 7

Problems

Northeastern

Department of Civil and Environmental Engineering
400 Snell Engineering
360 Huntington Ave.
Boston, MA 02115

CIVE 7381 Transportation Demand
Problem Set #2
Due: Wednesday, October 2, 2024

Problem 1
A random sample of students was drawn in which they reported the number of vehicular trips they made in the previous day. Let y_i = number of trips made by respondent i. The sample size was n = 100. y has a sample mean of 0.43. The (true) standard deviation of the population is not known but the sample standard deviation is equal to 0.84. The trip rate (average number of vehicular trips per day) is, as stated above, 0.43.

a) What is the standard error of the average trip rate?
b) Find the 95% confidence interval of the trip rate (since the sample size is 100 you can assume normal distribution)?
c) Suppose the results of this survey are being used to plan a later survey. How large should the sample be so that the precision of the trip rate, at the 95% confidence level, will be ±0.05?

Problem 2
You are given a set of aggregate data from 57 traffic analysis zones (TAZ) in the Chicago Area (the data is quite old as you can infer from the various variables that are included and the "official" description of the variables included in the table in the next page that reflect conditions at that time). For each of the 57 zones you have available the average trips per occupied dwelling unit, the average car ownership, the average household size, and three zonal social indices. The data is contained in the file PS2-2-data.xlsx. In a later homework you will use the data to develop, using linear regression, a model that predicts the trips/day (*TODU*) generated by a household. For now, you need to do a preliminary analysis of the data and understand the main trends and relationships.

a) Describe and briefly discuss the cause/effect relationships you think are relevant for trip generation rates.
b) Present relevant statistical summaries and descriptive statistics on the data provided. Check for the presence of outliers. At the minimum, your analysis should include summary statistics (e.g. median, mode, average, and standard deviation). For the relevant variables plot their histogram, using appropriate intervals. Develop the boxplots and identify any outliers. Comment on the shape of the distributions.
c) Examine relationships between pairs of the various variables. Use scatter plots to illustrate these relationships. Calculate the correlation coefficients among pairs of variables. Do the data support the relationships you propose in a)?

The data file contains the following variables:

Name	Description
TODU	*Trips per Occupied Dwelling Unit* Trips refer to the daily frequency of person-trips via motor vehicle (auto driver or passenger) or public transit made from a dwelling unit by members of that dwelling unit. All trips whose origins were other than "from home" were ignored.
ACO	*Average Car Ownership* Cars per dwelling unit.
AHS	*Average Household Size* Number of residents per dwelling unit.
SRI	*Job/Skills Rank Index* This index reflects two elements: (i) the proportion of blue-collar workers, defined as the ratio of craftsmen, operatives, and laborers to all employees; and (ii) educational level as measured by the proportion of persons 25 years and older completing eight or fewer years of schooling. The index attains a maximum value when no residents are in the blue-collar jobs category, and all adult residents have more than eight years of education
UI	*Urbanization Index* This index reflects three elements: (i) the ratio of children under five years of age to the female population of childbearing age; (ii) percentage of women who are in the labor force, and (iii) the percentage of single units to total dwelling units. The degree of urbanization index would be increased by (a) lower ratio in (i), (b) higher percentage in ii); and (c) lower proportion of single dwelling units. High values for this index imply less attachment to the home because of fewer children, higher likelihood of women being employed, and less permanency of dwelling unit type in terms of average tenure.
MI	*Minority Index* This index is defined as the proportion of an area's residents who are minorities.

Problem 3

Read the report **Summary of Travel Trends: 2022 National Household Travel Survey** and answer the following:

a) Comment on the main trends included in Section 3: *Household Travel* and include a graph or a table summarizing that you found most surprising and explain why you found it to be surprising.
b) Comment on the main trends included in Section 4: *Person Travel* and include a graph or a table that you found most surprising and explain why you found it to be surprising.
c) Repeat the same for Sections 5, 6, and 7.
d) What are the main emerging travel trends discussed in Section 9?

Solution to Problem 1

Problem 1a

Given a sample size of 100 students and a sample standard deviation of 0.84 trips, the standard error, based on **Equation 1**, is **0.084**.

$$SE = \frac{s}{\sqrt{n}}$$

Equation 1: Standard Error

Where:

SE = standard error
s = sample standard deviation
n = sample size

Problem 1b

At 95% confidence interval, $\alpha/2 = 0.5(100 - 95)\% = 2.5\%$. This implies that $z_{\alpha/2} = 1.96$. Using **Equation *2***, the confidence interval is from 0.265 trips to 0.595 trips or (0.430 ± 0.165) trips.

$$CI = \bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$$

Equation 2: Confidence Interval

Where:

CI = confidence interval
$\bar{x}$ = sample mean
$z_{\alpha/2}$ = confidence level value at $100(1-\alpha)\%$
s = sample standard deviation
n = sample size

Problem 1c

Using, **Equation *3***, the sample should have a size of 1,085 students in order to achieve a desired margin of error of ±0.05 at the 95% confidence level.

$$n = \left(\sigma\frac{z_{\alpha/2}}{\varepsilon}\right)^2$$

Equation 3: Number of Samples

Where:

n = sample size
σ = sample standard deviation
$z_{\alpha/2}$ = confidence level value at $100(1-\alpha)\%$
ε = desired margin of error

Problem 2

Problem 2a

Trip generation rates have historically been considered and proved to be a function of land use, certain household demographic characteristics, as well as some socio-economic factors.

Before analyzing the given data, I would expect the following:

- Higher average car ownership (ACO) will result in higher number of trips per occupied dwelling unit (TODU). For example, if ACO is higher, people may visit the grocery store more often than if ACO is lower.

- Average household size (AHS) can possibly be a good predictor of TODU. It is logical that the higher the AHS, the higher the TODU. For example, all other factors being equal, a household of size 6 is expected to make more trips than a household of size 2.

- The relationship between urbanization index (UI) TODU seems complex at first sight. High UI implies low ratio of children under 5 years to female population of childbearing age, high percentage of women in workforce, and high percentage of single units to total dwelling units. This means that fewer children will be going to school when UI is higher. Hence, it can be assumed that high UI leads to low TODU.

- At the most basic level, we can expect that higher minority index (MI) will result in lower TODU. Ceteris paribus, I will expect that minority groups are constrained (especially financially) that they cannot afford additional trips like ones for pleasure.

- All other factors being equal, I will expect that higher skills rank index (SRI), which means very few or no residents in the zone are working blue collar jobs and/or most of the residents have a high level of education, will result in low TODU because such people are more likely to work from home. But given that the data is quite old and work from home wasn't really a thing until the COVID-19 pandemic, that argument may not hold. Hence, one can expect that high SRI results in high TODU because residents typically don't work for very long hours and have a lot of free time than otherwise to embark on additional trips (e.g., for pleasure).

Problem 2b

The provided dataset was read into a Pandas DataFrame in Python. The descriptive statistics of the relevant variables have been illustrated in **DataFrame 1**. N.B. Only the zone column was excluded.

DataFrame 1: Descriptive Statistics for the Variables in the Dataset

```
relevant_variables = ['ACO', 'AHS', 'MI', 'SRI', 'UI', 'TODU']
dF2[relevant_variables].describe()
```

	ACO	AHS	MI	SRI	UI	TODU
count	57.000000	57.000000	57.000000	57.000000	57.000000	57.000000
mean	0.811754	3.185088	13.073158	49.560000	52.618772	5.373158
std	0.177655	0.389052	12.186774	15.844723	13.458622	1.325458
min	0.500000	1.830000	2.170000	20.890000	24.080000	3.020000
25%	0.670000	3.000000	6.820000	38.140000	44.800000	4.540000
50%	0.790000	3.190000	9.860000	49.370000	55.510000	5.100000
75%	0.920000	3.370000	15.080000	60.850000	61.090000	6.130000
max	1.320000	4.500000	62.530000	87.380000	83.660000	9.140000

The boxplots for the relevant variables, as defined in **DataFrame 1**, are illustrated in in **Figure *1***. From the boxplots, the variables **ACO**, **AHS**, **MI**, and **TODU** all contain some outliers, with the highest number of outliers observed in the variable **MI**, which is highly positively skewed.

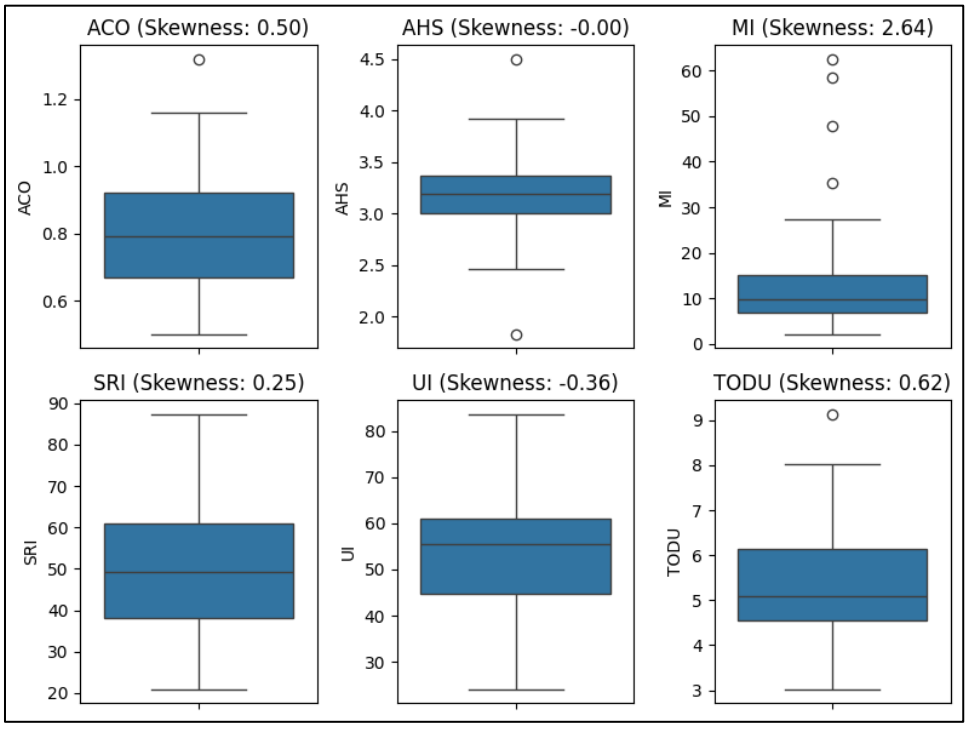

Figure 1: Boxplots of the Distributions of Relevant Variables

Histograms for the relevant variables have been illustrated in **Figure *2***.

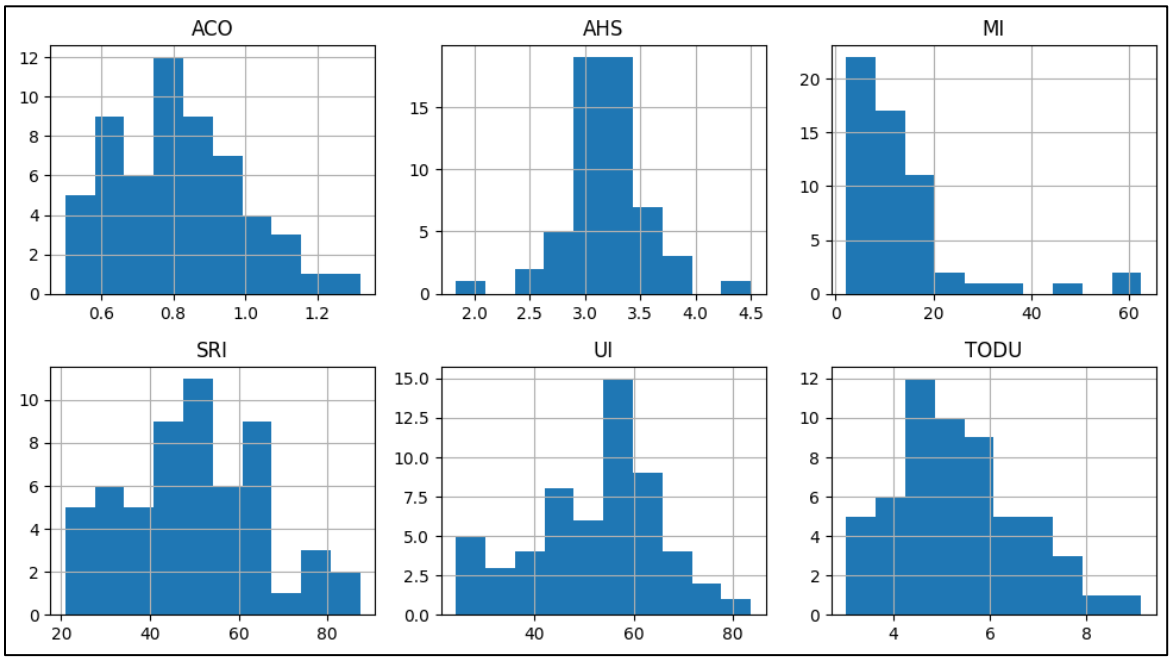

Figure 2: Histograms of the Distributions of the Relevant Variables

Problem 2c

The correlation and scatter matrices for the relevant variables has been illustrated in **Figure *3*** and **Figure *4***, respectively. They both confirm almost all the assumptions I made in Part a).

```
# Correlation matrix
correlation_matrix = dF2[relevant_variables].corr()
correlation_matrix
```

	ACO	AHS	MI	SRI	UI	TODU
ACO	1.000000	0.236115	-0.396768	0.316662	-0.713194	0.785194
AHS	0.236115	1.000000	0.084910	-0.517667	-0.571897	0.374812
MI	-0.396768	0.084910	1.000000	-0.180489	0.248613	-0.246000
SRI	0.316662	-0.517667	-0.180489	1.000000	0.020830	0.186341
UI	-0.713194	-0.571897	0.248613	0.020830	1.000000	-0.756952
TODU	0.785194	0.374812	-0.246000	0.186341	-0.756952	1.000000

Figure 3: Correlation Matrix of Relevant Variables

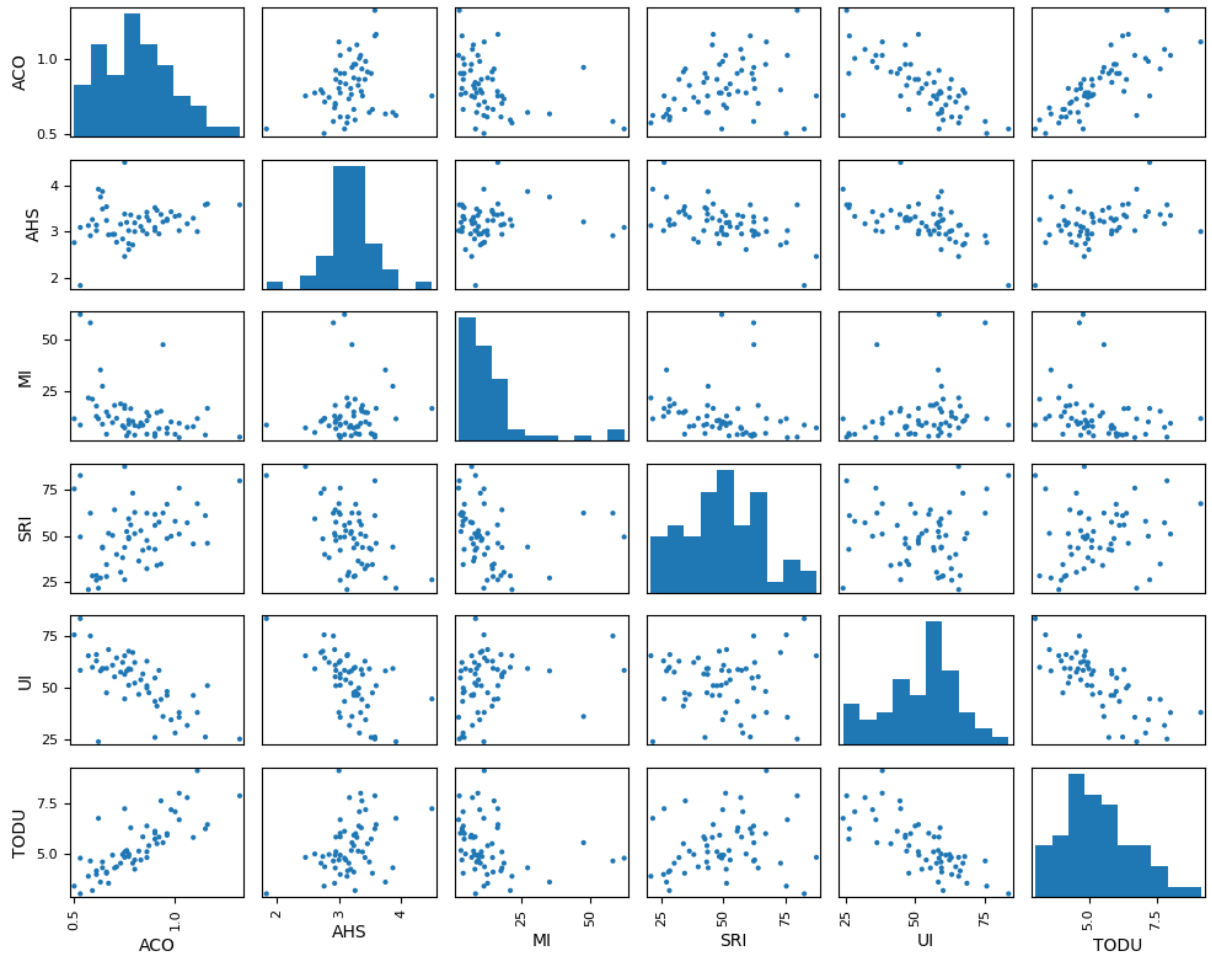

Figure 4: Scatter Matrix of Relevant Variables

Problem 3

Comments on **Summary of Travel Trends: 2022 National Household Travel Survey** by the Federal Highway Administration, an agency within the United States Department of Transportation.

Section 3: Household Travel

The main trend in household travel is the significant decline in work-related trips, shopping trips, school trips, and social and recreational trips; be it vehicle trips, person trips person miles traveled (PMT) or vehicle miles traveled (VMT). On average, the annual person trips from 2017 to 2022 dropped by roughly 37% (from 3,140 to 1,990 person trips). This decrease shows how much the pandemic has affected travel behavior in the United States.

To me, what I found most surprising is the sharp decline for PMT per household in shopping trips (a decrease of 53%) from 2017 to 2022. I know the pandemic affected the way Americans shop but I didn't expect this sharp decline. Another surprising trend to me was that while there were fewer trips made in 2022, the average length of each trip was longer compared to 2017 and previous NHTS years (**Table *1***), which explains that Americans are now more willing to travel longer or many have moved to stay further away from the city, in the suburbs.

Table 1: Average Person Trip Length by Trip Purpose

Category	Survey Year	Trip Purpose							
		To/ From Work	Work-Related Business	Shopping	Other Family/ Personal Errands	School/ Church	Social and Recreational	Other	All
Trends based on self-reported distance	1983	8.5	21.8	5.4	7.3	4.9	12.3	8.2	8.7
	1990	10.7	28.2	5.4	8.6	5.4	13.2	10.3	9.5
	1995	11.6	20.3	6.1	7.6	6.0	11.3	22.8	9.1
	2001	12.1	28.3	7.0	7.8	6.0	11.4	43.1	10.0
	2009	11.8	20.0	6.5	7.0	6.3	10.7	51.5	9.7
Trends based on network-calculated distance	2009 (adj)	11.0	18.3	5.9	6.3	5.8	9.7	48.9	8.9
	2017	11.5	25.9	7.1	7.1	6.4	10.4	49.1	10.7
	2022	13.4	20.7	5.8	8.6	6.2	17.1	54.1	12.6
MOEs	2009 (adj)	0.3	1.9	0.2	0.3	0.3	0.6	14.3	0.3
	2017	0.3	6.4	0.5	0.3	1.2	0.5	7.3	0.4
	2022	1.3	6.5	0.5	0.9	0.8	2.7	24.0	1.0

Section 4: Person Travel

The trends in person travel (Section 4) are very similar to the trends in household travel (Section 3) in terms of decreased trip activity. To wit, total person trips also fell sharply compared to 2017, mirroring household travel trends.

I think that the increase in social and recreational trips is surprising given the overall trend of reduced travel (**Figure *5***). This shows how Americans adjusted their travel patterns to prioritize leisure activities, likely as part of pandemic recovery.

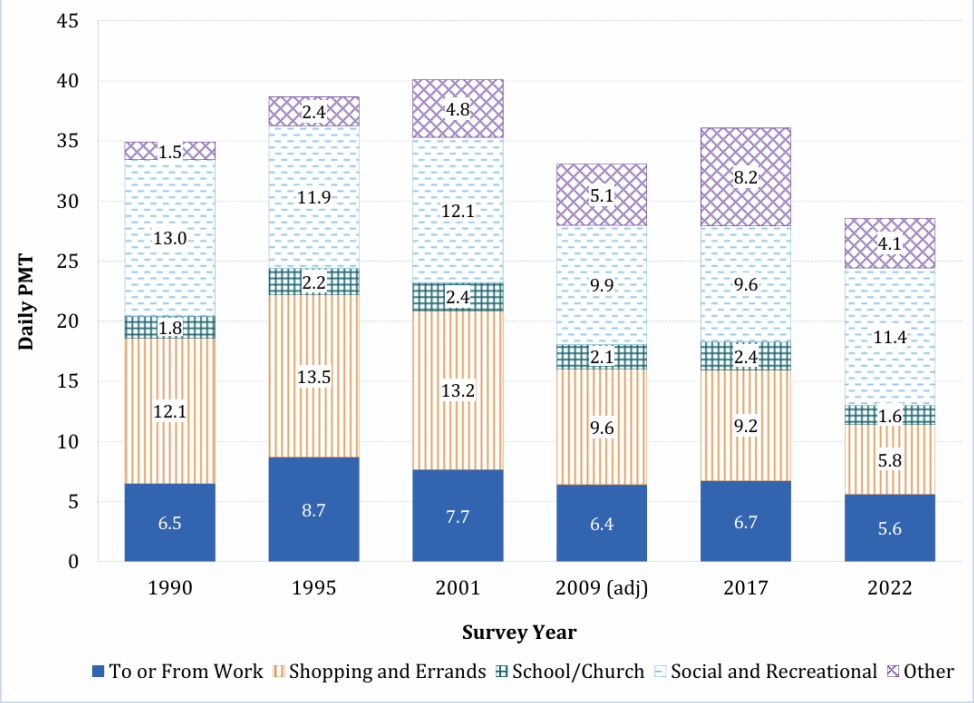

Figure 5: Daily PMT per Person by Trip Purpose

Section 5: Private Vehicle Travel

The main trend here is that although private vehicle travel remained the dominant travel mode in the United States in 2022, as compared to other modes like public transit, the amount of time spent driving decreased. There was a 36% reduction in time spent driving across all drivers compared to 2017.

Actually, nothing under this section emerged particularly surprising to me, as the results are consistent with what was obtained for Section 3 and Section 4. What nearly surprised me was that from 1990 through 2017, the vehicle occupancy estimates, measured as person miles per vehicle mile, have stayed about the same (**Table *2***). However, there was a decline across all trip purposes from 2017 to 2022.

Table 2: Average Vehicle Occupancy for Selected Trip Purposes

Category	Survey Year	Trip Purpose				
		To/From Work	Shopping	Other Family/ Personal Errands	Social and Recreational	All
Average vehicle occupancy	1990 (adj)	1.14	1.71	1.84	2.08	1.64
	1995	1.14	1.74	1.78	2.04	1.59
	2001	1.14	1.79	1.83	2.03	1.63
	2009	1.13	1.78	1.84	2.20	1.67
	2017	1.18	1.82	1.82	2.10	1.67
	2022	1.08	1.53	1.60	1.99	1.52
MOEs	2009	0.01	0.05	0.04	0.06	0.03
	2017	0.01	0.05	0.13	0.04	0.04
	2022	0.04	0.09	0.07	0.19	0.08

Note: The "Other Family/Personal Errands" category includes trips such as to the post office, dry cleaners, or library. The "All" category includes other trip purposes not shown, such as trips to school, church, doctor, dentist, and work-related business trips. For explanations of adjustments as well as specific differences in survey methods over time, please refer to Section 1.2.

Section 6: Vehicle Use and Availability

Here, I see that households in 2022 owned slightly fewer vehicles on average than in previous surveys. Households without a vehicle increased slightly compared to 2017, and the percentage of households with three or more vehicles decreased. What surprised me was that out of the 128 million households in the United States, about 10.7 million were without a vehicle (**Table *3***).

Table 3: Number of Households by Availability of Household Vehicles

Category	Survey Year	Household Vehicle Availability					Average Vehicles per Household
		0 Vehicles	1 Vehicle	2 Vehicles	3+ Vehicles	All	
Number of Households (thousands)	1969	12,876	30,252	16,501	2,875	62,504	1.16
	1977	11,538	26,092	25,942	11,840	75,412	1.59
	1983	11,548	28,780	28,632	16,411	85,371	1.68
	1990	8,573	30,654	35,872	18,248	93,347	1.77
	1995	7,989	32,064	40,024	18,914	98,990	1.78
	2001	8,716	33,757	39,938	24,955	107,365	1.89
	2009	9,828	36,509	41,077	25,688	113,101	1.86
	2017	10,567	39,648	39,125	28,869	118,208	1.88
	2022	10,712	42,271	48,347	26,214	127,545	1.83
MOEs	2009	49	302	274	270	0	—
	2017	0	0	272	272	0	—
	2022	1,197	1,905	1,921	1,567	2,383	—

Section 7: Commute Travel Patterns

Main trend is that in 2022, there was a notable increase in the number of workers who teleworked several days a week, while the overall number of commute trips dropped, as expected. What surprised me was that commute travel times have increased slightly, indicating that those who are still commuting tend to live further from their workplaces (**Table *4***).

Table 4: Commute Patterns by Mode of Transportation

Category	Survey Year	Commute Details		
		Average Commute Trip Length (Miles)	Average Commute Travel Time (Minutes)	Average Commute Speed (Miles per Hour)
All modes	2009 (adj)	11.03	—	25.79
	2017	11.46	26.58	23.42
	2022	13.43	27.72	25.51
All modes MOEs	2009 (adj)	0.27	0.31	0.31
	2017	0.34	0.56	0.28
	2022	1.30	1.96	0.92
Privately owned vehicle	2009 (adj)	11.26	—	27.01
	2017	11.84	25.01	25.22
	2022	13.56	26.94	26.59
Privately owned vehicle MOEs	2009 (adj)	0.23	—	0.31
	2017	0.38	0.56	0.33
	2022	1.25	1.90	0.89
Public transit	2009 (adj)	10.18	—	11.42
	2017	12.09	58.11	11.63
	2022	7.53	43.05	10.68
Public transit MOEs	2009 (adj)	1.54	—	1.01
	2017	1.15	2.06	0.73
	2022	2.50	4.60	3.81
Walk	2009 (adj)	0.98	—	4.76
	2017	1.19	15.26	3.15
	2022	1.28	25.05	2.80
Walk MOEs	2009 (adj)	0.23	—	0.51
	2017	0.73	1.59	0.18
	2022	0.53	6.71	0.41

Section 9.2: Emerging Travel Trends

I identified a few. There are more people now more than ever teleworking and engaging in online shopping, there's an increased use in micromobility especially in the urban areas, and lastly, there's a decline in public transportation use.